Slide 1

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha
UMass Amherst

Detmar Meurers, Ramon Ziai
Universität Tübingen

CALICO, Amherst
June 10, 2010

UMASS AMHERST

Eberhard Karls
Universität Tübingen

Developing and Integrating ICALL Systems
Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction
Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning
Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation
Summary
References

Slide 2

Overview

► Background: TAGARELA, a Portuguese ICALL system

► Integration of ICALL into different pedagogical settings
  ► Individualized instruction
  ► Distance learning

► Development of TAGARELA processing architecture
  ► flexible integration of NLP modules to support diverse activity demands
  ► prioritization of feedback depending on activity and learner
  ► improved feedback by supporting multiple perspectives on learner data

Developing and Integrating ICALL Systems
Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction
Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning
Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation
Summary
References

Slide 3

Introducing TAGARELA

► TAGARELA is an intelligent workbook for Portuguese we created at The Ohio State University (Amaral & Meurers 2006; Amaral 2007; Amaral, Meurers & Silva 2006).

► It addresses real-life needs identified in interviews with OSU foreign language instructors (Amaral 2004).

► TAGARELA has six different activity types
  ► Listening Comprehension
  ► Reading Comprehension
  ► (Picture) Description
  ► Fill-in-the-Blanks
  ► Rephrasing
  ► Vocabulary

Developing and Integrating ICALL Systems
Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction
Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning
Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation
Summary
References

Slide 4

Developing and Integrating ICALL Systems
Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction
Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning
Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation
Summary
References

Slide 1 (5/30)

How TAGARELA got started

- ▸ The Ohio State University has the longest running individualized language instruction (I.I.) center in the US.

- ▸ In 2004, we developed an individualized program for Portuguese (Silva & Amaral 2005).

- ▸ This provided an interesting opportunity to explore the integration of a web-based tutoring system.

- ▸ In 2005, a two-year project was funded to develop the TAGARELA system (Amaral, Meurers & Silva 2006).

UMASS AMHERST

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Slide 2 (6/30)

Original pedagogical context
Individualized Instruction Program (IIP)

- ▸ General characteristics of Individualized Instruction:
  - ▸ self-paced, mastery-based program
  - ▸ student uses material for self-study
  - ▸ work with the instructor is done by appointment
  - ▸ feedback on exercises provided on student's request

- ▸ Meetings with the instructor are used to:
  - ▸ ask specific questions about the material
  - ▸ grade compositions and exams
  - ▸ take oral exams

- ▸ Consequences:
  - ▸ students receive less feedback on their performance
  - ▸ students are less exposed to the target language
  - ▸ it is harder for instructors to monitor students' performance

UMASS AMHERST

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Slide 3 (7/30)

Original pedagogical context
TAGARELA in individualized instruction

- ▸ We have designed TAGARELA to fill some of the pedagogical shortcomings of the IIP setup.
  - ▸ TAGARELA offers on the spot individualized feedback on spelling, morphological, syntactic and semantic errors.
  - ▸ It provides opportunities for students to practice their listening, reading, and writing skills.

- ▸ TAGARELA was designed to be an independent web-based workbook.

- ▸ It presented a sequence of activities that could be incorporated into any pedagogical material.

- ▸ Students use menus to navigate through its six exercise types at different levels.

UMASS AMHERST

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Slide 4 (8/30)

UMASS AMHERST

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

New pedagogical context
Distance learning program

- The University of Massachusetts offers more than 1,500 courses online.

- In 2009, we received a grant from the College of Humanities and Fine Arts to develop a Portuguese online language sequence.

- In the current summer semester, two elementary level courses are being offered for the first time.

New pedagogical context
Distance learning compared to individualized instruction

- *Different* from individualized instruction, in distance learning
  - the student's work is not self-paced,
  - work with instructor is not done through individual appointments,
  - meetings with instructors are computer mediated.

- *Similar* to individualized instruction, in distance learning
  - most materials are used for self-study,
  - students do not have the classroom exposure to L2,
  - it is harder for instructors to monitor students' performance,
  - students receive less feedback on their individual performance.

Motivations for using ICALL in distance learning

- Students receive immediate feedback on written exercises.
  - This adresses the clear need for more feedback in distance learning, also identified in two CALICO 2010 talks this morning:
    - Ana Gimeno
    - Wesley Curtis, Victoria Russell and Nancy Blain

- It helps instructors monitor students' performance.

- It provides extra support for self-study.

- It expands the contexts in which L2 is used by learners.

- It is realized in the predominant online medium, and it can easily be incorporated into sequences of activities.

TAGARELA for distance learning

- TAGARELA is integrated into the UMass course management system.

- It does not have a separate interface or menu system.

- The TAGARELA activities appear as part of a learning module in the course management system.

- The activities are selected to complement the sequence of activities of the course.

TAGARELA as an independent workbook (Live Demo)

Development of the NLP architecture

Three issues we addressed to advance the computational side

1. To provide individualized feedback, the different activities make use of different NLP modules.
   - ‣ Implement an architecture supporting flexible integration of different NLP modules.

2. Can feedback be prioritized depending on
   - ‣ the nature of the activity (activity model), and
   - ‣ the history of the learner (learner model)?

3. Can we support multiple perspectives on learner data (linguistic analysis vs. learner conceptualization) to improve feedback?

Background: Architecture was redesigned in Java (Ziai 2009).

Flexibly handling a range of activities

‣ Activities differ with respect to the nature of
   - ‣ the student input
   - ‣ the NLP required for that input
   - ‣ the feedback to be generated

Activity	Input	NLP	Feedback prioritizes
FIB	Words	Spelling, lexicon	Word form, missing word
Vocabulary	Phrase	same as FIB	Missing+Extra word, word form
Rephrasing	Sentence	same as FIB + parsing	Form-related errors
Description	Sentence	same as Rephrasing	Missing word, agreement
Reading	Sentence	same as Rephrasing + content	Meaning-related errors
Listening	Sentence	same as Reading	Meaning-related errors

‣ How can such a heterogeneous set of activities and their demands be handled by the same system?

UIMA-based architecture

‣ UIMA: Unstructured Information Management Architecture (Ferrucci & Lally 2004)
   - ‣ each NLP tool annotates the input
   - ‣ UIMA data repository is common to all components (Götz & Suhre 2004)

‣ Crucial difference to older NLP architectures:
   - ‣ input is *enriched*, not *transformed*
   - ‣ similar to stand-off XML in manual corpus annotation

‣ Data structures for linguistic annotation ("type system") are modeled explicitly
   - ‣ supporting a demand-driven integration of NLP components
   - ‣ specific to the needs of the activity

The need to prioritize feedback

- ▶ Where a learner sentence contains multiple errors, feedback should be provided one at a time (Heift 2003).
- ▶ How should the feedback messages be prioritized?
  - a) It depends on the activity.
  - b) It depends on the learner.

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

17 / 30

Feedback prioritization through activity model

- ▶ The original TAGARELA always provided feedback on meaning errors before feedback on form errors.
- ▶ Yet some activities target form aspects of language.
- ➡ Support explicit prioritization in activity model.
  - ▶ Rephrasing activities: *form* < *meaning*
  - ▶ Reading comprehension: *meaning* < *form*

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

18 / 30

Feedback prioritization through activity model
Form-focused feedback for Rephrasing

1. Meaning-related error: omission of *Unidos* ('United')
2. Form-related error: use of *das* (fem.plur) instead of *dos* (masc.plur) ('from the')
➡ System gives feedback on **second**, form-related error

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

19 / 30

Feedback prioritization through activity model
Meaning-focused feedback for Reading Comprehension

1. Meaning-related error: *é* ('is') instead of *tem* ('has')
2. Form-related error: *ano* ('year') does not agree in number with *quinze* ('fifteen')
➡ System gives feedback on **first** error

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

20 / 30

Feedback prioritization through learner model

- ▸ The learner model in TAGARELA records for each learner:
  - ▸ number and nature of activities completed
  - ▸ number and nature of errors detected
    - ▸ Only *one* error is selected for feedback, but *all* errors are recorded in the learner model.
- ▸ Learner model serves as additional ranking criterion for feedback where learner answers contain multiple errors.
- ▸ Example: Vocabulary activity
  - ▸ Student 1 has a record of problems with lexical choice.
  - ▸ Student 2 has a record of problems with spelling.

Feedback prioritization through learner model

Example student 1: Feedback on lexical choice

Feedback prioritization through learner model

Example student 2: Feedback on spelling error

Improving feedback through annotation-based learner language representation

- ▸ Certain Portuguese words are syntactically complex.
- ▸ Contraction: preposition + determiner/pronoun
  - ▸ *no* = *em* (in) + *o* (the)
  - ▸ *nela* = *em* (in) + *ela* (it)
  - ▸ *destes* = *de* (of) + *estes* (these)
  - ▸ *às* = *a* (to) + *as* (the)
- ▸ Encliticization:
  - ▸ *comprá-lo* = *comprar* (to buy) + *o* (it)
  - ▸ *compram-nas* = *compram* (buy$_{3.pl.pres}$) + *as* (them)
  - ▸ *comprei-a* = *comprei* (bought$_{1.sg.past}$) + *a* (it)

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

25 / 30

Problematic feedback

Regiões do Brasil

O Brasil está política e geograficamente dividido em cinco regiões. Os limites que cada região (Norte, Nordeste, Sudeste, Sul e Centro-Oeste) coincidem sempre com as fronteiras dos estados que as compõem.

A região Norte ocupa a maior parte do território brasileiro, com uma área que corresponde a 45,27% da área total do País. Formada por sete Estados, tem sua área quase totalmente dominada pela bacia do Rio Amazonas.

A região Nordeste pode ser considerada a mais heterogênea do País. Dividida em quatro grandes zonas – meio-norte, zona da mata, agreste e sertão –, ocupa 18,26% do território nacional e tem nove estados.

O Sudeste é formado por quatro Estados. Está é a região de maior importância econômica do País, onde está concentrado também o maior índice populacional - 42,63% dos brasileiros.

Já o Sul, região mais fria do País, com ocorrências de geadas e neve, é a que apresenta menor área, ocupando 6,75% do território brasileiro e com apenas três Estados. Os rios que cortam sua área formam a bacia do Paraná e ocupam toda sua totalidade e são de grande importância para o País, principalmente pelo seu potencial hidrelétrico.

Finalmente, a região Centro-Oeste tem sua área dominada basicamente pelo Planalto Central Brasileiro e pode ser dividida em três porções: maciço goiano-mato-grossense, bacia de sedimentação do Paraná e as depressões. Ela é formada por quatro Estados e nela está a capital do Brasil.

Questão 2

Em que região fica o Rio Amazonas?

O Amazonas fica no região norte.

à á â ã é ê í ó ô õ ú ç
Á Á Â Ã É Ê Í Ó Ô Õ Ú Ç   Enviar

Questões: 1 2 3 4 5 6 7
Próxima Questão (3)

Análise:

Input: O Amazonas fica no região norte.

There is an agreement error in *gender* between the *determiner* and the noun in the sequence *o região norte* from your answer.

To see a possible answer, click here.

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

26 / 30

Problematic feedback explained

Questão 2

Questões: 1 2 3 4 5 6 7
Próxima Questão (3)

Em que região fica o Rio Amazonas?

O Amazonas fica no região norte.

à á â ã é ê í ó ô õ ú ç
Á Á Â Ã É Ê Í Ó Ô Õ Ú Ç   Enviar

Análise:

Input: O Amazonas fica no região norte.

There is an agreement error in *gender* between the *determiner* and the noun in the sequence *o região norte* from your answer.

To see a possible answer, click here.

Report Errors & Suggestions

▸ System's interpretation: *no* = *em* (in) + *o* (the)

  ▸ syntactically analyzed: $[_{PP}$ em $[_{NP}$ o$_{masc}$, região$_{fem}$, norte$]]$
  → Agreement error between *o* and *região*.

▸ Student's interpretation of feedback:

  ▸ There is no "*o região norte*" in the sentence I wrote.

⇒ Feedback should refer to forms entered by the student
(Amaral & Meurers 2009).

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
Learner data representation

Summary

References

27 / 30

Addressing the issue

▸ The system needs to connect the surface form provided by the student with the system analysis of this input.

▸ An annotation-based NLP architecture (→ UIMA) readily supports this with multiple parallel layers of annotation for the learner input.

▸ The tokenization mismatch can be addressed by representing both surface and deep tokenizations of the learner input, and the mapping between the two.

  ▸ Refer to surface form when generating the feedback.

Developing and Integrating ICALL Systems

Luiz Amaral, Flávia Cunha, Detmar Meurers, Ramon Ziai

Introduction

Integrating into new Pedagogical Setting
Background: TAGARELA
Individualized Instruction
TAGARELA for Ind. Instruction
Distance Learning
TAGARELA for Dist. Learning

Development of NLP Architecture
Handling a range of activities
UIMA architecture
Prioritizing Feedback
Impact of activity models
Impact of learner models
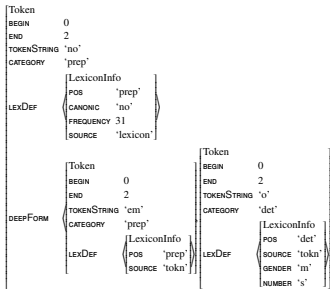Learner data representation

Summary

References

28 / 30

Solution: Model complex tokens in UIMA

Result: Helpful feedback

Questões: 1 2 3 4 5 6 7
Questão 2   Próxima Questão (3)

Em que região fica o Rio Amazonas?

O Amazonas fica no região norte.

à á â ã é ê í ó ô õ ú ç
À Á Â Ã E É Í Ó Ô Õ Ú Ç

Enviar

Análise:

Input: **O Amazonas fica no região norte.**

There is an agreement error in **gender** between the **determiner** and the **noun phrase** in the sequence **no região norte** for your answer.

To see a possible answer, click here.

Report Errors & Suggestions

Summary

- We discussed different pedagogical settings and their consequences for using the intelligent language tutoring system TAGARELA in those settings.
  - Individualized Instruction
  - Distance Learning

- We motivated new UIMA-based architecture capable of
  - handling heterogeneous NLP needs of diverse activities
  - representing multiple perspectives on learner data

- We discussed the new capability of the system to prioritize feedback based on activity and learner models.
  - Which prioritization strategy based on the information now stored in the activity and learner models works best, is an open empirical question → SLA study

---

References

Amaral, L. (2004). ICALL Specifications. Ms. Department of Linguistics & Department of Spanish and Portuguese. Ohio State University.

Amaral, L. (2007). Designing Intelligent Language Tutoring Systems: integrating Natural Language Processing technology into foreign language teaching. Ph.D. thesis, The Ohio State University.

Amaral, L. (2010). Revisiting Current Paradigms in Computer Assisted Language Learning Research and Development. *Ilha do Desterro* Brazil.

Amaral, L. & D. Meurers (2006). Where does ICALL Fit into Foreign Language Teaching? URL http://purl.org/net/icall/handouts/calico06-amaral-meurers.pdf. 23rd Annual Conference of the Computer Assisted Language Instruction Consortium (CALICO), May 19, 2006. University of Hawaii.

Amaral, L. & D. Meurers (2008). From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context: Extending the Conceptualization of Student Models for Intelligent Computer-Assisted Language Learning. *Computer-Assisted Language Learning* 21(4), 323–338. URL http://purl.org/dm/papers/amaral-meurers-call08.html.

Amaral, L. & D. Meurers (2009). Little Things With Big Effects: On the Identification and Interpretation of Tokens for Error Diagnosis in ICALL. *CALICO Journal* 27(1). URL http://purl.org/dm/papers/amaral-meurers-09.html.

Amaral, L., D. Meurers & G. Silva (2006). Using Intelligent Computer Assisted Language Learning Systems to Support Portuguese Instruction. APSA Conference. Oct. 5–7, 2006.

Amaral, L., D. Meurers & R. Ziai (to appear). Analyzing Learner Language: Towards A Flexible NLP Architecture for Intelligent Language Tutors. *Computer-Assisted Language Learning* URL http://purl.org/dm/papers/amaral-meurers-ziai-10.html.

Ferrucci, D. & A. Lally (2004). UIMA: an architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering* 10(3–4), 327–348.

Götz, T. & O. Suhre (2004). Design and implementation of the UIMA Common Analysis System. *IBM Systems Journal* 43(3), 476–489.

Heift, T. (2003). Multiple Learner Errors and Meaningful Feedback: A Challenge for ICALL Systems. *CALICO Journal* 20(3), 533–548. URL http://purl.org/calico/heift-03.html.

Silva, G. & L. Amaral (2005). Challenges for Establishing a Portuguese Individualized Instruction Program in the US. SAMLA Conference. November 5, 2005. Atlanta–GO.

Ziai, R. (2009). A Flexible Annotation-Based Architecture for Intelligent Language Tutoring Systems. Master's thesis, Universität Tübingen, Seminar für Sprachwissenschaft.